

IB/2004/052842

# 证 明

本证明之附件是向本局提交的下列专利申请副本

申 请 日： 2003.12.30 ✓

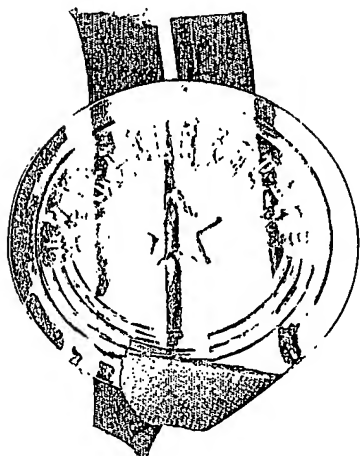
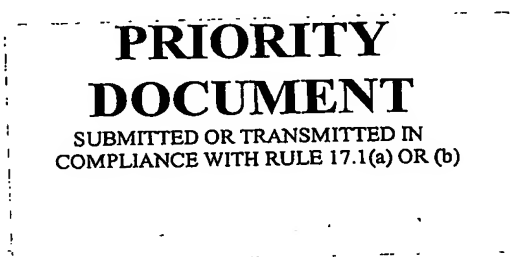
申 请 号： 2003101245205 ✓

申 请 类 别： 发明

发明创造名称： 一种X M L 数据的压缩与解压缩方法及装置

申 请 人： 皇家飞利浦电子股份有限公司

发明人或设计人： 安东尼·摩里尔



中华人民共和国  
国家知识产权局局长

王 荣 川

2004 年 10 月 10 日

## 权利要求书

1. 一种 XML 数据的压缩方法, 包括步骤:
  - a. 接收 XML 数据;
  - b. 对 XML 数据进行编码;
  - c. 对编码后的 XML 数据进行块打包;
  - d. 插入指示数据到块打包后的 XML 数据块之间, 以得到压缩后的 XML 数据, 该指示数据用于识别特定的数据。
2. 如权利要求 1 所述的方法, 其中所述的指示数据是位于一个空数据块中。
3. 如权利要求 2 所述的方法, 其中所述的指示数据是位于一个空数据块的块首。
4. 一种 XML 数据的压缩方法, 包括步骤:
  - a. 接收 XML 数据;
  - b. 插入指示数据到 XML 数据中, 该指示数据用于识别特定的数据;
  - c. 对含有指示数据的 XML 数据进行压缩, 以得到压缩后的 XML 数据。
5. 如权利要求 4 所述的方法, 其中步骤 b 包括步骤:

对所述的 XML 数据进行分析, 以得到一组无用的数据作为指示数据标志;

在特定数量的指示数据标志后插入相应的指示数据;

将剩下的指示数据标志以另一组无用的数据代替。
6. 如权利要求 4 所述的方法, 其中步骤 b 包括步骤:

对所述的 XML 数据进行分析, 以得到一组无用的数据;

转换特定数量的所述无用数据为指示数据包;

将所述的指示数据放入所述的指示数据包。

7. 如权利要求 5 或 6 所述的方法, 所述无用的数据是下述数据之一: 制表标记, 空格标记和回车标记。

8. 一种压缩后的 XML 数据的解压缩方法, 包括步骤:

- a. 接收压缩后的 XML 数据, 该压缩后的 XML 数据含有指示数据;
- b. 解压缩该压缩后的 XML 数据, 其中该步骤包括步骤 (i) 获得所述的指示数据;
- c. 根据该指示数据将解压缩后的对应的 XML 数据丢弃。

9. 如权利要求 8 所述的方法, 其中所述的指示数据是位于一个空数据块中。

10. 如权利要求 8 所述的解压缩方法, 其中步骤 b 中的步骤 i 包括步骤:

对所述的压缩后的 XML 数据进行块首解码, 从而找到一个空数据块;  
从该空数据块的块首中获得该指示数据。

11. 如权利要求 8 所述的解压缩方法, 还包括步骤:

根据一个特定的条件修正该指示数据的内容, 其中步骤 c 根据修正后的指示数据内容来执行。

12. 如权利要求 8 所述的解压缩方法, 所述丢弃的 XML 数据是对应于所述压缩后的 XML 数据中特定的数据块。

13. 一种压缩后的 XML 数据的解压缩方法, 包括步骤:

- a. 解压缩该压缩后的 XML 数据, 以得到解压缩后的 XML 数据;
- b. 从所述的解压缩后的 XML 数据获得一个指示数据, 该指示数据用于识别特定的数据;
- c. 根据该指示数据将对应的解压缩后的 XML 数据丢弃。

14. 如权利要求 13 所述的解压缩方法, 其中所述的指示数据是插入于原始的 XML 数据中的。

15. 如权利要求 13 所述的解压缩方法, 其中步骤 b 包括步骤:

在所述的 XML 数据找到一个指示数据标志;

根据该指示数据标志获得该指示数据。

16. 如权利要求 13 所述的解压缩方法, 还包括步骤:

根据一个特定的条件修正该指示数据的内容, 其中步骤 c 根据修正后的指示数据内容来执行。

17. 一种 XML 数据的压缩装置, 包括:

一个接收装置, 用于接收 XML 数据;

一个编码装置, 用于对 XML 数据进行编码;

一个块打包装置, 用于对编码后的 XML 数据进行块打包;

一个插入指示数据块装置, 用于插入指示数据到块打包后的 XML 数据之间, 以得到压缩后的 XML 数据, 该指示数据用于识别特定的数据。

18. 如权利要求 17 所述的装置, 所述的指示数据是位于一个空数据块中。

19. 一种 XML 数据的压缩装置, 包括:

一个接收装置, 用于接收 XML 数据;

一个插入指示数据包装置, 用于将指示数据插入到 XML 数据中, 该指示数据用于识别特定的数据;

一个压缩装置, 用于对插入指示数据的 XML 数据进行压缩, 以得到压缩后的 XML 数据。

20. 如权利要求 19 所述的装置, 其中所述的插入指示数据包装置包括:

一个定位装置, 用于对所述的 XML 数据进行分析, 以得到一组无用的数据作为指示数据标志;

一个数据插入装置, 用于在特定数量的指示数据标志后插入相应的指示数据, 并将剩下的指示数据标志以另一组无用的数据代替。

21. 如权利要求 20 所述的装置, 所述无用的数据是下述数据之一: 制表标记, 空格标记和回车标记。

22. 一种压缩后的 XML 数据的解压缩装置, 包括:

- 一个接收装置, 用于接收压缩后的 XML 数据, 该压缩后的 XML 数据含有指示数据;
- 一个数据处理装置, 用于解压缩该压缩后的 XML 数据, 并获得所述的指示数据;
- 一个丢弃装置, 用于根据该指示数据将该压缩后的对应的 XML 数据丢弃。

23. 如权利要求 22 所述的装置, 其中所述的指示数据是位于一个空数据块中。

24. 如权利要求 22 所述的装置, 其中所述的数据处理装置包括:

- 一个空数据块检测装置, 用于对所述的压缩后的 XML 数据进行块首解码, 从而找到一个空数据块;
- 一个指示数据获得装置, 用于从该空数据块的块首中获得该指示数据。

25. 如权利要求 22 所述的装置, 还包括一个分析器, 用于根据一个特定的条件修正该指示数据的内容, 其中所述的丢弃装置根据修正后的指示数据内容来执行。

26. 如权利要求 24 所述的装置, 其中所述的指示数据是插入于原始的 XML 数据中。

27. 如权利要求 24 所述的装置, 其中所述的指示数据是从解压缩后的 XML 数据中获得的。

28. 如权利要求 24 所述的装置, 所述的数据处理装置包括一个检测提取装置, 用于在解压缩后的 XML 数据中找到一组指示数据标志, 并根据该指示数据标志获得该指示数据。

# 说明书

## 一种 XML 数据的压缩与解压缩方法及装置

### 背景技术

本发明涉及一种数据压缩与解压缩的方法和装置，尤其涉及一种 XML (Extensible Markup Language, 可扩展的升级语言) 数据压缩与解压缩的方法和装置。

XML 是一种文本格式，它在数据交换中正越来越普遍。越来越多的标准，例如：多媒体领域、MPEG-7 和 TV-Anytime，都使用 XML 文本格式表示数据。

XML 是一种冗长的格式，即 XML 表示数据和结构的方式导致一个相对较大的文本。因此，为了传输或者储存，需要仔细考虑数据压缩。最普通的压缩方法是 Zlib，如大家都熟悉的 zip (.zip 文件) 和 gzip (.gz 文件)，它是基于霍夫曼、LZ77 或两者。

现有技术中，压缩装置对 XML 数据进行压缩，然后将压缩的 XML 数据传送给解压缩装置，解压缩装置对压缩的 XML 数据进行解压并进行分析。

图 1 是现有技术的一个压缩器的结构示意图。压缩器 100 (compressor) 包括 LZ77 编码器 102、霍夫曼编码器 104 和块打包器 106 (block wrapping)，压缩器 100 根据 Zlib 格式对 XML 数据进行压缩。

压缩器 100 首先接收 XML 数据，LZ77 编码器 102 根据 LZ77 算法对 XML 数据进行编码，产生一束代码字 (codeword) 和字符 (literals)。该字符包含来自 XML 数据中不能压缩的字节，一个代码字可以将 XML 数据中以前就已经遇到过的，也就是多余的数据转换成一序列的字节。典型的代码字包含长度和间距，该长度是以前遇到过序列的长度，间距是从字节中序列的开端到当前字节的间距。

霍夫曼编码器 104 对代码字和字符进行霍夫曼编码，输出一序列不同长度的代码，并产生一个霍夫曼列表。

块打包器 106 从霍夫曼编码器 104 得到霍夫曼列表，将数据打包成块，每个块可以使用不同的霍夫曼列表或者甚至完全不用 LZ77 编码和霍夫曼编码。这里打包有 3 种可能性：旁路压缩、使用默认的霍夫曼列表、使用常规的霍夫曼列表，这 3 种可能性是基于如实际的压缩率和平均信息量。每个块都以一个块首 (block header) 开始。最后输出压缩的 XML 数据，

并发送给解压缩装置。

图 2 是现有技术的一个解压缩装置中解压缩器和分析器的结构示意图。解压器 (decompressor) 200 对压缩的 XML 数据进行解压, 获得 XML 数据。解压器 200 包括块首解码器 (block header decoding) 202、霍夫曼解码器 (Huffman decoding) 204 和 LZ77 解码器 (LZ77 decoding) 206。

块首解码器 202 对压缩的 XML 数据进行解码, 获得一个霍夫曼列表及不同长度的代码和/或字符, 霍夫曼解码器 204 再对压缩的 XML 数据进行解码, 获得代码字和字符, 最后输入 LZ77 解码器 206 进行解码, 获得 XML 数据。

分析器 210 带有用于 XML 数据的简单应用程序接口 (SAX), 用于对 XML 数据进行 SAX 分析, 获得事件类型 (Event\_Type) 和事件数据 (Event\_Data)。这里的 SAX 实际上是对 XML 数据处理的标准, 它非常简单, 所以非常快。SAX 按序处理 XML 数据, 所以与基于 Zlib 的按序解压器 200 非常匹配。SAX 是基于事件的概念, 事件是为在按序处理 XML 数据期间, SAX 分析遇到的实体而产生的。通过出现的事件类型来指示分析器 210 事件的类型, 这样分析器 210 可以相应地对事件数据进行分析处理, 得到分析后的 XML 数据。

这里在 SAX 分析之前, 系统只是将 XML 数据作为一序列的字符 (即压缩器不假定数据的特性)。而在 SAX 分析之后, 不同的 XML 实体例如元素和非元素 (字符) 就被区分开来。因此, SAX 分析后的输出并不包含单个的字符, 而是一序列的事件, 每个事件对应一个实体, 这个实体由 XML 数据中许多不同的字符所构成。

由于现有技术中, 从一个大的压缩文件中重获特殊的数据对于接收机来说是一种负担, 但在大 XML 数据上完成压缩比在小 XML 数据上压缩要好, 尤其在带宽昂贵 (如广播) 的域中, 压缩效率最佳化是非常重要的。进一步, 如果目标接收机没有存储, 那么不可能在一个数据库中以一个解压格式保存全部的数据, 它最多将数据以压缩格式保存或者一直等到数据再次传送。所以现有技术中带有大量资源, 如大存储能力, 的装置不能直接对大 XML 文件进行操作。且带有有限资源, 如小存储能力的装置更不能以解压格式或者数据库格式储存数据, 它们只能够在压缩文件的基础上重新获得数据。

## 发明内容

本发明针对现有技术中存在的问题，提供了一种用于 XML 数据压缩和解压缩方法和装置。

本发明提供了一种 XML 数据的压缩的方法，首先，接收 XML 数据，对 XML 数据进行编码；然后，将编码后的 XML 数据处理打包成数个数据块；最后，在所述数个数据块之间插入指示数据，得到压缩的 XML 数据，该指示数据用于识别特定的数据。

本发明提供了另一种 XML 数据的压缩的方法，首先，接收 XML 数据；然后，插入指示数据到 XML 数据中，该指示数据用于识别特定的数据；最后，对含有指示数据的 XML 数据进行压缩，以得到压缩后的 XML 数据。

本发明提供了一种 XML 数据的解压缩的方法，首先，接收压缩后的 XML 数据，该压缩后的 XML 数据含有指示数据；然后，解压缩该压缩后的 XML 数据，并在该解压缩过程中获得所述的指示数据；最后，根据该指示数据将解压缩后的对应的 XML 数据丢弃。

本发明提供了另一种 XML 数据的解压缩的方法，首先，解压缩该压缩后的 XML 数据，以得到解压缩后的 XML 数据；然后，从所述的解压缩后的 XML 数据获得一个指示数据，该指示数据用于识别特定的数据；最后，根据该指示数据将对应的解压缩后的 XML 数据丢弃。

本发明通过避免分析 XML 数据中不相关的数据，从而可以加速分析的过程，这样就使接收器的工作加速；由于只对 XML 数据中相关部分进行处理，所以就可以处理较大的 XML 数据。而所有要传输的 XML 信息就可以被分割成较大 XML 数据中的一小块数据，这要远比处理小 XML 数据中一大块的数据好的多，因为前者使用 Zlib 进行压缩比后者要好的多，所以可节约带宽。

通过参照结合附图所进行的如下描述和权力要求，本发明的其它目的和成就将是显而易见的，并对本发明也会有更为全面的理解。



### 附图说明

本发明通过实例的方式，参照附图进行详尽的解释，其中：

图 1 是现有技术的一个压缩器的结构示意图；

图 2 是现有技术的一个解压缩装置中解压缩器和分析器的结构示意图；

图 3 是根据本发明的一个实施例的压缩装置的结构框图；

图 4 是根据本发明的一个实施例的压缩方法的流程图；

图 5 是根据本发明的一个实施例的解压缩装置的结构示意图；

图 6 是根据本发明的一个实施例的解压缩方法的流程图；

图 7 是根据本发明的另一个实施例的压缩装置的结构框图；

图 8 是根据本发明的另一个实施例的压缩方法的流程图；

图 9 是根据本发明的另一个实施例的解压缩装置的结构框图；

图 10 是根据本发明的另一个实施例的解压缩方法的流程图；

在所有的附图中，相同的参照数字表示相似的或相同的特征和功能。

### 具体实施方式

图 3 是根据本发明一个实施例的压缩器结构框图，该压缩器 100 包括一个 LZ77 编码器 102，一个霍夫曼编码器 104，一个块打包器 106 和一个插入指示数据块装置 302。

所述 LZ77 编码器 102，用于对 XML 数据进行 LZ77 编码，还可作为接收装置，用于接收 XML 数据。霍夫曼编码器 104，用于对经过 LZ77 编码的 XML 数据进行霍夫曼编码，同时提供霍夫曼列表。LZ77 编码器 102 和霍夫曼编码器 104 可以一起构成编码装置，用于对 XML 数据进行编码。

块打包器 106，用于根据霍夫曼列表将经过霍夫曼编码后的 XML 数据打包成数个数据块，每个数据块的块首有部分霍夫曼列表。

插入指示数据块装置 302，用于根据霍夫曼列表将指示数据插入所述数据块之间，以得到压缩后的 XML 数据，该指示数据用于识别特定的数据。所述的指示数据是位于一个空数据块中。

图 4 是根据本发明一个实施例的压缩方法的流程图，首先接收 XML 数据（步骤 S402），例如，接收的 XML 数据为：

<Entry><Word>Aback</Word><Definition>saldiufhcnw</Definition></Entry>.....

然后对 XML 数据进行编码，包括进行 LZ77 编码（步骤 S404）和进行霍夫曼编码（步骤 S406）。XML 数据经过 LZ77 编码（步骤 S404）后，得到一束代码字（codeword）和字符（literals），这里的代码字就是 XML 数据中重复的字符“Word>”，它的长度为 5，它的距离即第一个“Word>”到下一个“Word>”之间的间隔为 12。字符就是其他不能被压缩的字符，例如“Aback”等等。

对 XML 数据进行霍夫曼编码（步骤 S406），获得不同长度的代码，同时产生霍夫曼列表。例如：20 个字符‘E’‘n’‘t’‘r’‘y’‘>’‘<’‘W’‘o’‘r’‘d’‘>’‘A’‘b’‘a’‘c’‘k’‘<’‘/’，经过霍夫曼编码后得到 20 个不同长度的代码为十六进制的：6C 75 9E A4 A2 A9 6E 6C 87 9F A2 94 6E 71 92 91 93 9B 6C 5F。

根据霍夫曼列表将经过霍夫曼编码的 XML 数据进行块打包，打包成数个数据块（步骤 S408）。例如将以字母‘A’开头的单词打包在一个数据块，将以字母‘B’开头的单词打包在下一个数据块，依次排序，得到数个数据块。

插入指示数据到块打包后的 XML 数据块之间（步骤 S410），以得到压缩后的 XML 数据（步骤 S412），该指示数据用于识别特定的数据，这里特定的数据是所需要的数据，例如单词‘car’。

所述的指示数据是位于一个空数据块中，且指示数据是位于一个空数据块的块首。

压缩后的 XML 数据如表 1 所示：

数据块号	块首 (Header)	内容 (Contents)
0		6C 75 9E A4 A2 A9 6E 6C 87 9F A2 94 6E
1 (指示数据块)	霍夫曼列表 '0' C '1' End of Block	空
2		"Aback</[...]" = 71 92 91 93 9B 6C 5F ...
3 (指示数据块)	霍夫曼列表 '0' E '1' End of Block	Empty
4		"Car</[...]" = ...
...	...	...

表 1

从表 1 可以看出, 数据块 0 包含的内容与编码后的 XML 数据“<Entry><word>”相对应, 即 6C 75 9E A4 A2 A9 6E 6C 87 9F A2 94 6E; 数据块 1 即指示数据块的块首插有指示数据 ‘C’, 该数据块是空数据块, 没有任何数据; 数据块 2 和数据块 3 与数据块 0、1 相似。数据块 4 是以字母 ‘C’ 开头的单词, 该数据块的内容就是与单词 “Car” 相对应的字符, 即和前述 “6C 75 ” 等等相似的字符。

图 5 是根据本发明一个实施例解压缩装置的结构示意图, 该装置解压缩装置包括一个解压缩器 500, 一个有限状态机 (finite state machine, FSM) 510, 一个检测指示数据块装置 508 和一个分析器 512。

解压缩器 500 还包括一个块首解码器 502, 一个霍夫曼解码器 204 和一个 LZ77 解码器 206。

块首解码器 502 用于对压缩后的 XML 数据块进行块首解码, 在进行块首解码时每遇到一个新的数据块, 会产生一个数据块信号, 并将该信号发送给有限状态机 510。块首解码器 502 还用于找到一个空数据块, 并将该空数据块提供给检测指示数据块装置 508。块首解码器 502 还用于产生霍夫曼列表, 同时还可以作为接收装置, 接收压缩后的 XML 数据。

霍夫曼解码器 204, 用于根据霍夫曼列表对经过块首解码后的压缩后的 XML 数据进行解码。

LZ77 解码器 206, 用于对压缩后的 XML 数据进行 LZ77 解码, 获得 XML 数据。该压缩后的 XML 数据含有指示数据。

检测指示数据块装置 508, 用于从块首解码器 502 提供来的空数据块的块首中获得指示数据, 并发送给分析器 512。所述解压缩器 500 和检测指示数据块装置 508 一起构成一个数据处理装置, 用于解压缩该压缩后的 XML 数据。

分析器 512 根据一个特定的条件修正该指示数据的内容, 产生相应的跳跃信号, 并发送给有限状态机 510, 该特定的条件对应于分析器 512 的一个特定的应用, 即分析器 512 需要的数据, 例如单词 ‘car’。该指示数据的修正可有两种结果, 一种为执行该指示数据的内容, 即相应的跳跃信号为要求有限状态机 510 丢弃一些不相关的数据; 另一种为略过该指示数据, 即相应的跳跃信号内容为空。

有限状态机 510 根据数据块信号以及修正后的指示数据内容, 即跳跃信号来丢弃该压缩后的对应的 XML 数据。该分析器 512 和有限状态机 510 一起构成丢弃装置, 用于根据该指示数据将该压缩后的对应的 XML 数据丢弃。

图 6 是根据本发明一个实施例的解压缩方法的流程图，首先接收压缩的 XML 数据(步骤 S602)，该压缩的 XML 数据包含指示数据块。

然后解压缩该压缩的 XML 数据，包括：

对该压缩的 XML 数据进行块首解码（步骤 S604），从而找到一个空数据块，并产生数据块信号，例如对该数据块 1 进行块首解码，就产生数据块 1 的数据块信号。

对指示数据块进行检测（步骤 S606），如检测到指示数据块，例如对数据块 1 的内容进行块首解码，得知该数据块是空数据块，那么说明该数据块是指示数据块，则就从数据块 1 的块首获得指示数据的内容（步骤 S610），例如 ‘C’。

如在步骤 S606 中未检测到指示数据块，则对下一个数据块即数据块 2 进行检测，得知数据块 2 不是指示数据块，那么就对其进行霍夫曼解码（步骤 S612），再进行 LZ77 解码（步骤 S614），获得数据块 2 的数据。

然后，根据指示数据的内容和分析器的内部状态，即一个特定的条件，来判断是否要产生跳跃信号（步骤 S616），亦即根据一个特定的条件修正该指示数据的内容。该特定的条件是一个特定的应用，即分析器的内部状态需要的数据，例如单词 ‘car’，然后根据指示数据 ‘C’，对指示数据的内容进行修正，即产生一个跳跃信号，要求直接跳到 “C” 部分。

接下来，根据数据块信号和跳跃信号丢弃不相关的数据块（步骤 S618），例如正在寻找单词 “Car”，那么就判断出 “Car” 是以字母 ‘C’ 开头后的单词，出现在后面的数据块中，所以产生跳越信号，将不相关的数据块，即直到数据块 3 的数据块信号出现前的所有数据块 2 的数据（“B” 部分）丢弃。由于解压缩后的 XML 数据不是块结构，故需根据数据块信号来控制每一个被丢弃的数据块。

同理，按照前述方法，从数据块 3 的块首获得指示数据内容 ‘E’ (610)，同时获得数据块 4 的数据（步骤 S614），然后根据指示数据 ‘E’ 和正在寻找的单词 “Car” 进行判断（步骤 S616），由于单词 “Car” 是在以字母 ‘E’ 开头的单词之前，所以，就不产生跳越信号，然后对相关的数据块即数据块 4 进行分析（步骤 S620），最后获得分析后的 XML 数据，例如单词 “Car”。

这里对解压缩后的对应的 XML 数据进行丢弃，是根据修正后的指示数据内容即跳跃信号来执行的。

如果在步骤 S616 中的判断结果为否定，说明不需要进行丢弃，则，直接对相关的数据块进行分析（步骤 S620），并获得分析后的 XML 数据（步骤 S622）。

图 7 是根据本发明另一个实施例的压缩装置的结构示意图，该压缩装置包括一个分析器 702，一个压缩器 100。

分析器 702 还包括一个定位装置 704，用于得到一组无用的数据作为指示数据标识，同时还作为接收装置，接收 XML 数据；一个数据插入装置，用于在特定数量的指示数据标识后插入相应的指示数据，将剩下的指示数据标志以另一组无用的数据代替。无用的数据是下述数据之一：制表标记、空格标记、回车标记等等。

压缩器 100，用于对插入指示数据的 XML 数据进行压缩，得到压缩的 XML 数据。

图 8 是根据本发明另一个实施例的压缩方法的流程图，首先接收 XML 数据(步骤 S802)，例如 XML 数据是：

```
<Entry><Word>➔Aback</Word><Definition>saldiufhcnw</Definition></Entry>...
```

```
<Entry><Word>➔Car</Word><Definition>lzidnuvgrvgs</Definition></Entry>...
```

然后对 XML 数据进行 SAX 分析，找到 XML 数据中一组没用的字符，例如可以是一组 20 个 ‘➔’（制表标记），还可以是空格标记、回车标记等等。将这一组没用的字符 ‘➔’ 作为指示数据标识（步骤 S806）。

将特定数量的指示数据标识 ‘➔’，如 14 个，后插入指示数据（步骤 S808），例如 ‘C’。再将剩下的 ‘➔’ 用其他没用的数据替代（步骤 S809），例如用空格替代。获得的 XML 数据为：

```
<Entry><Word>➔<!--C-->Aback</Word><Definition>saldiufhcnw</Definition></Entry>...
```

```
<Entry><Word>➔<!--E-->Car</Word><Definition>lzidnuvgrvgs</Definition></Entry>...
```

这里，还可以对 XML 数据进行分析，以得到一组无用的数据，例如 ‘➔’（制表标记）；再将特定数量的无用数据转换为指示数据包；将所述的指示数据放入所述的指示数据包，获得的 XML 数据如上所述。

然后再对包含指示数据的 XML 数据进行压缩，也就是对包含指示数据的 XML 数据进行 LZ77 编码（步骤 S810）；对经过 LZ77 编码的 XML 数据进行霍夫曼编码（步骤 S812）；将经过霍夫曼编码的 XML 数据打包成数个数据块（步骤 S814）；最后获得压缩的 XML 数据（步骤 S816）。

这里所述指示数据和数据块标识在 XML 数据被压缩以前插入 XML 数据。这里插入的指示数据和数据块标识对于解压缩装置是明显的，也就是解压缩装置将利用它们来忽略一些数据，这样就使解压缩装置的功能更加强大。

图 9 是根据本发明另一个实施例中解压缩装置的示意图, 该解压缩装置包括一个解压缩器 200, 一个检测提取装置 904, 一个有限状态机 510 和一个分析器 512。

解压缩器 200, 用于对压缩后的 XML 数据进行解压, 该压缩后的 XML 数据含有指示数据, 其中指示数据是插入于原始的 XML 数据中, 同时解压缩器 200 作为接收装置, 接收压缩后的 XML 数据。

检测提取装置 904, 用于在解压缩后的 XML 数据中找到一组指示数据标识, 根据该指示数据标识获得该指示数据, 并将该指示数据发送给分析器 512, 同时检测提取装置 904 产生指示数据标识信号, 并将该指示数据标识信号发送给有限状态机 510。解压缩器 200 和检测提取装置 904 一起可构成数据处理装置。

分析器 512 根据一个特定的条件修正该指示数据的内容, 该特定的条件是一个特定的应用, 即分析器 512 所需要的数据。然后对该指示数据的内容进行修正, 产生相应的跳跃信号, 并发送给有限状态机 510。

有限状态机 510 根据指示数据标识信号以及修正后的指示数据内容即跳跃信号, 来丢弃该压缩后的对应的 XML 数据。该分析器 512 和有限状态机 510 一起构成丢弃装置, 用于根据该指示数据将该压缩后的对应的 XML 数据丢弃。

图 10 是根据本发明另一个实施例的解压缩方法的流程图, 首先接收压缩的 XML 数据(步骤 S1002), 然后对压缩后的 XML 数据进行解压缩(步骤 S1004), 得到解压缩后的 XML 数据。

从所述的解压缩后的 XML 数据获得一个指示数据, 该指示数据用于识别特定的数据, 具体步骤如下:

对 XML 数据中的指示数据标识, 例如 “→”, 进行检测(步骤 S1006), 如果检测到, 那么就产生指示数据标识信号(步骤 S1008)。

提取该数据块标识后的指示数据(步骤 S1009), 例如 “C”。

然后, 根据指示数据的内容和分析器的内部状态, 即一个特定的条件, 来判断是否要产生跳跃信号(步骤 S1010), 即为根据一个特定的条件修正该指示数据的内容。也就是根据所述指示数据 “C” 以及一个特定应用, 即分析器的内部状态所需要的数据, 判断是否产生一个跳跃信号(skip signal)。例如正在寻找单词 ‘car’, 那么就判断出 “Car” 是以字母 ‘C’ 开头后的单词, 出现在后面的数据块中, 所以产生跳越信号, 要求将不相关的数据丢弃。

接下来, 如果在步骤 S1010 中产生一个要求丢弃数据的跳跃信号, 则根据数据块信号和

跳跃信号丢弃不相关的数据块（步骤 S1012），即将直到下一个指示数据标识信号出现前的所有数据丢弃，并回到步骤 S1006 继续进行检测判断。

同理，按照前述方法在检测到下一个数据块标识时，即下一个“→”，就获得其后面的指示数据内容‘E’（步骤 S1009）。然后根据所述指示数据“C”以及一个特定应用，即分析器的内部状态所需要的数据，判断是否产生一个跳跃信号（skip signal）（步骤 S1010）。例如正在寻找单词‘car’，那么就判断出“Car”是以字母‘E’开头后的单词之前，所以就不产生跳越信号，然后就对相关的 XML 数据块进行分析（步骤 S1014），最后获得分析后的 XML 数据（步骤 S1016），例如单词‘car’。

这里对应的解压缩后的 XML 数据丢弃，是根据修正后的指示数据内容，即跳跃信号来执行丢弃的。

如果在步骤 S1006 或步骤 S1010 的判断结果是否定的，则直接对相关的数据块进行分析（步骤 S1014），并获得分析后的 XML 数据（步骤 S1016）。

从本发明实施例中可以看出，通过避免分析 XML 输入数据中不相关的数据块，从而可以加速分析的过程，这样就使在接收端的工作加速；由于只对 XML 数据中相关部分进行处理，那么就可以处理较大的 XML 数据输入；所有要传输的 XML 信息就可以被分割成大 XML 数据中的一小块数据，这要远比处理小 XML 数据中一大块的数据好的多，因为前者使用 Zlib 进行压缩比后者要好的多，所以可节约带宽。

本发明由于对较大 XML 输入数据进行压缩，所以可以有更好的压缩。由于解压缩装置不必等待信息的重传，所以在解压缩装置中的存储器中压缩的 XML 数据可以提供对信息较快的访问。

本发明插入指示数据可与现有的压缩标准/方案兼容，从而使得压缩后的 XML 数据与现有的解压缩装置相兼容。

本发明将指示数据和 XML 数据作为一体，所以指示数据总是可以与 XML 数据的内容相匹配，即使在内容更新的情况下也是如此。本发明也不需要另外单独给指示数据一个传输信道，这就节约了在一个单独信道传输数据带来的多余开销，而且插入 XML 数据时，指示数据也是被 Zlib 进行压缩。

虽然经过对本发明结合具体实施例进行描述，对于在本技术领域熟练的人士，根据上文的叙述作出的许多替代、修改与变化是显而易见的。因此，当这样的替代、修改和变化落入附后的权利要求的精神和范围之内时，应该被包括在本发明中。

# 说明书附图

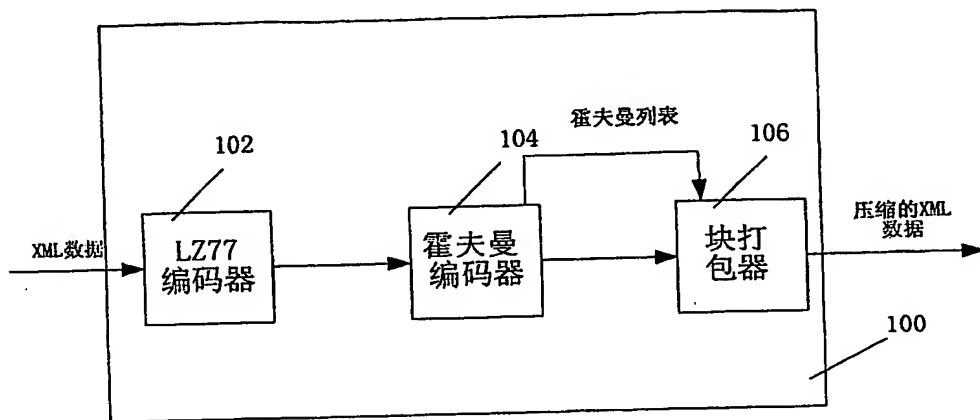


图1

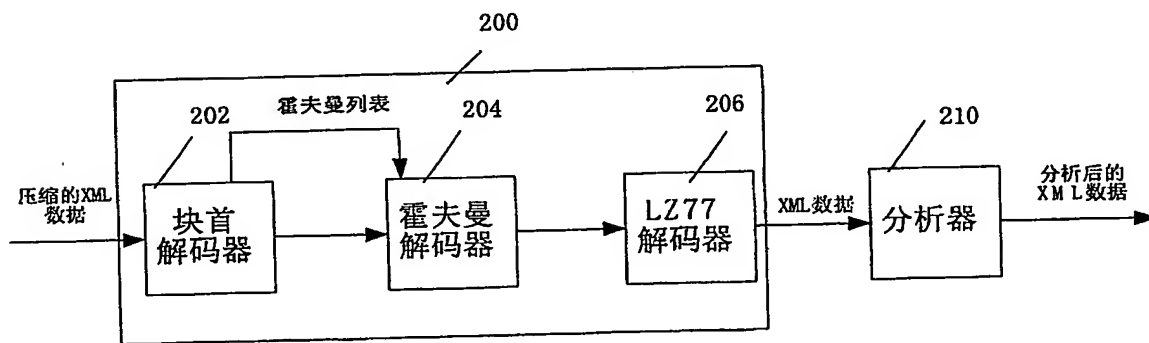


图2

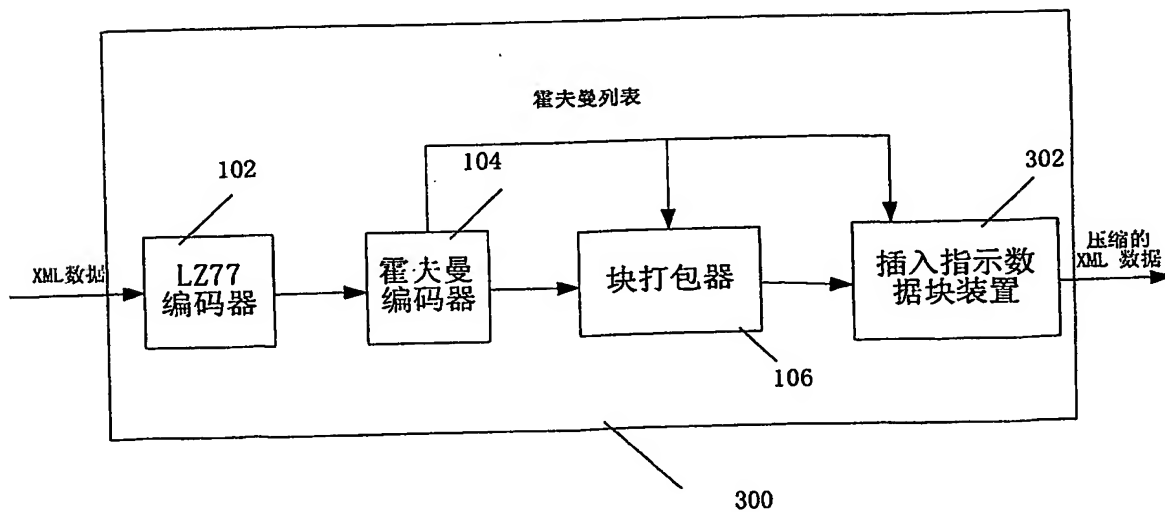


图3



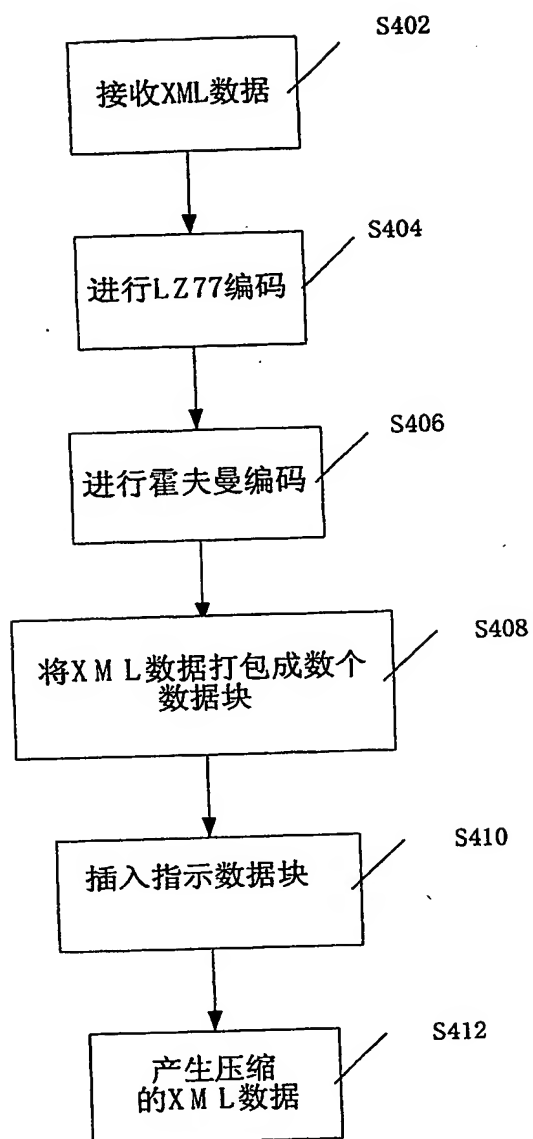


图4

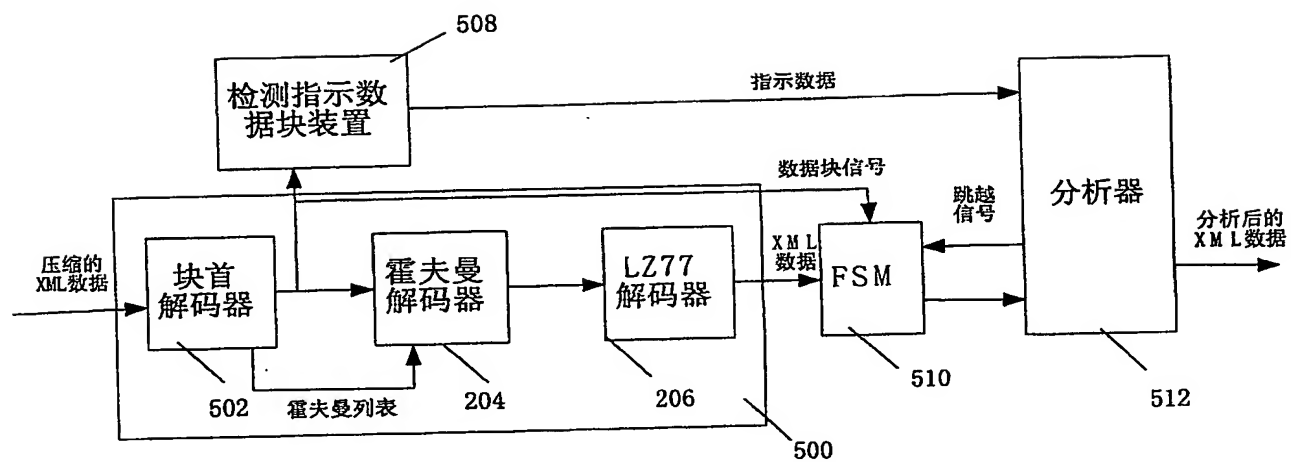


图5

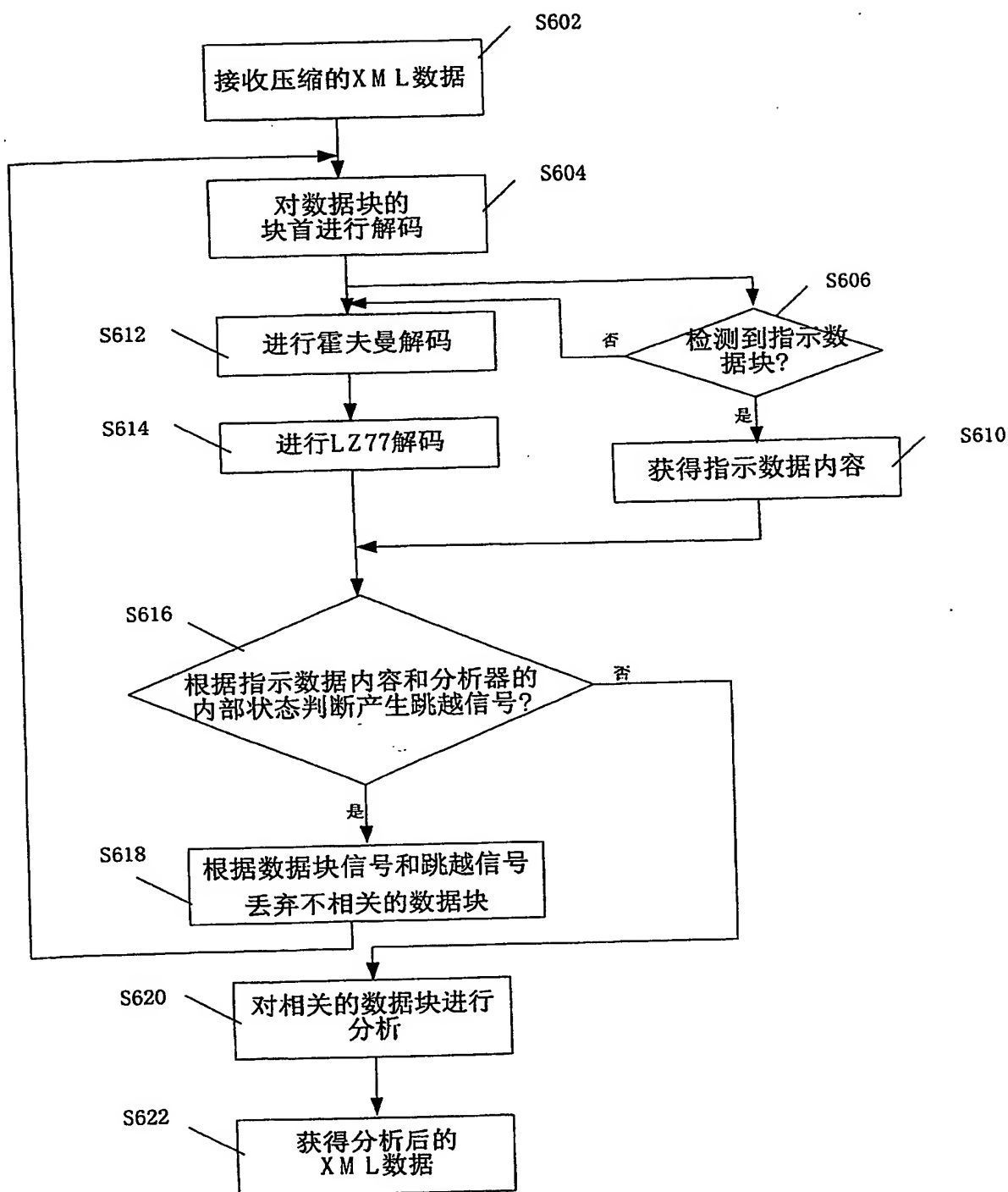


图6

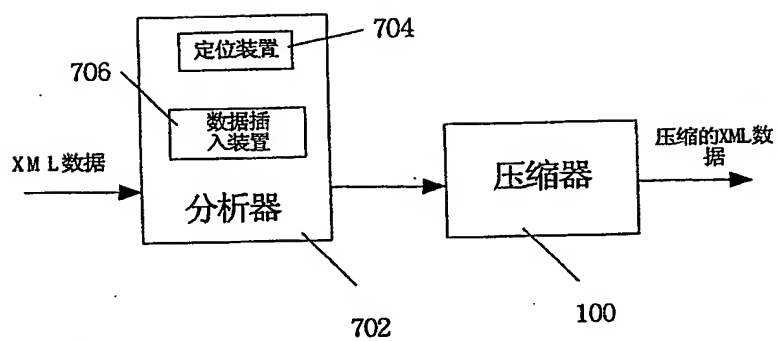


图7

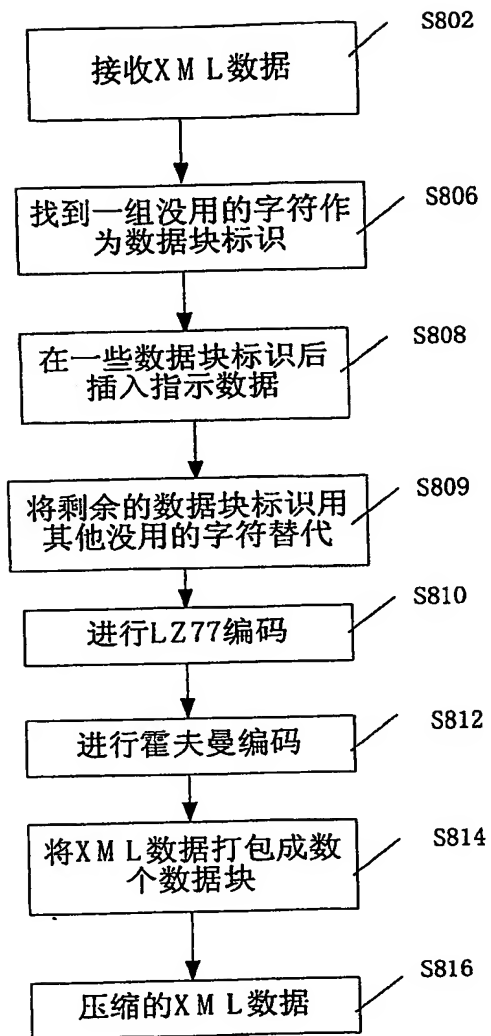


图8

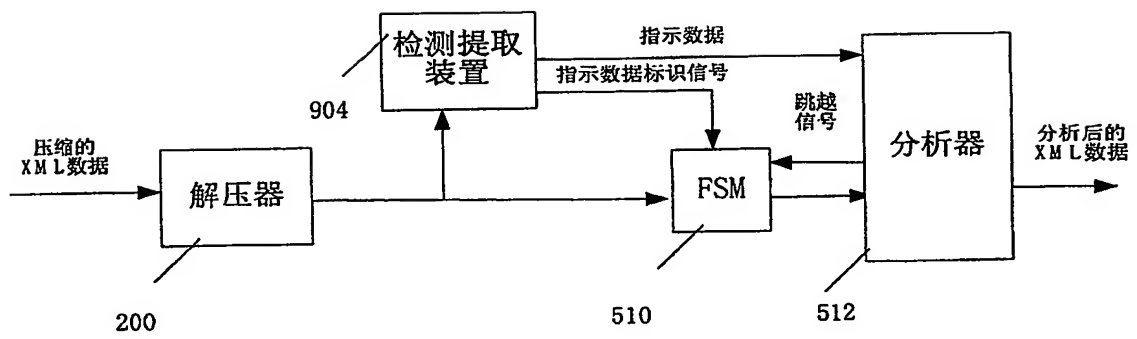


图9

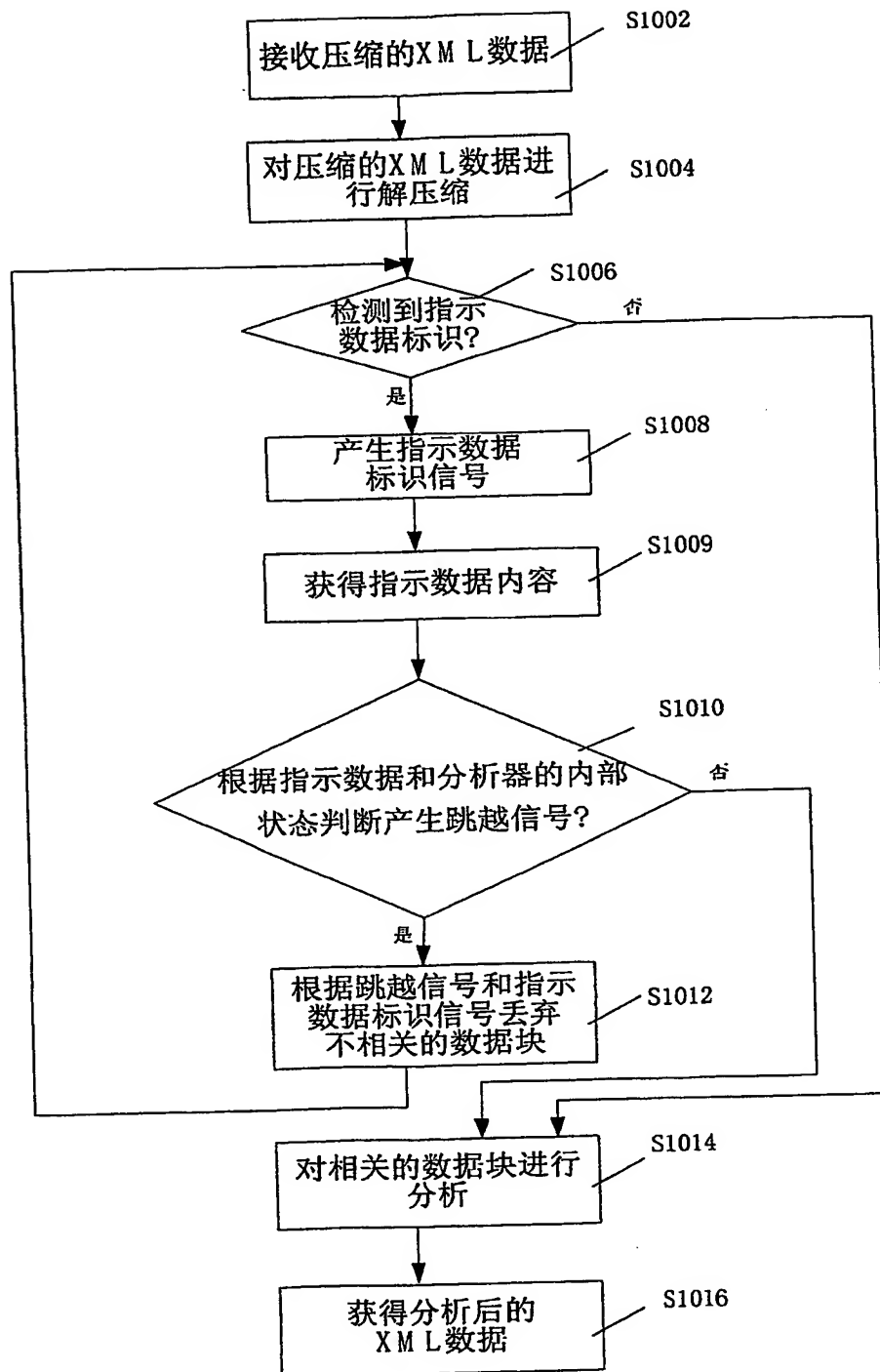


图10

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record.**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☒ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**